
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

# Machine learning models for covid-19 future forecasting

Ramesh Kumar Mojjada [a,*], Arvind Yadav [a], A.V. Prabhu [b], Yuvaraj Natarajan [c]

[a] Department of Computer Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, India
[b] Department of Electronics and Communication Engineering Koneru Lakshmaiah Education Foundation, Guntur
[c] ICT Academy, Chennai, Tamil Nadu, India

ARTICLE INFO

*Article history:*
Received 5 October 2020
Received in revised form 20 October 2020
Accepted 24 October 2020
Available online xxxx

*Keywords:*
COVID-19
Exponential process of smoothing
Future forecasting
R2 score adjusted
Machine learning supervised

ABSTRACT

Computational methods for machine learning (ML) have shown their meaning for the projection of potential results for informed decisions. Machine learning algorithms have been applied for a long time in many applications requiring the detection of adverse risk factors. This study shows the ability to predict the number of individuals who are affected by the COVID-19[1] as a potential threat to human beings by ML modelling. In this analysis, the risk factors of COVID-19 were exponential smoothing (ES). The Lower Absolute Reductor and Selection Operator, (LASSo), Vector Assistance (SVM), four normal potential forecasts, such as Linear Regression (LR)). [2] Each of these machine-learning models has three distinct kinds of predictions: the number of newly infected COVID 19 people, mortality rates and the recovered COVID-19 estimates in the next 10 days. These approaches are better used in the latest COVID-19 situation, as shown by the findings of the analysis. The LR, that is effective in predicting new cases of corona, death numbers and recovery.

Selection and peer-review under responsibility of the scientific committee of the Emerging Trends in Materials Science, Technology and Engineering.

## 1. Introduction

Over the last decade, machine learning (ML) proved to be an important study field, addressing many incredibly complicated and elaborate problems in the real world. [Nearly all real-world areas, including healthcare, autonomous vehicles (AV) [3,4], Enterprises software, NLP, smart robotics, sports, climatic simulation, voicing and image processing [6,7] were included in the application fields.

Important ML areas are expected. Many standard ML algorithms were used to direct potential outcomes in many application fields, including temperature, disease forecasting, stock market prediction, and disease prediction. Different models of regression and neural networks are broadly applicable to forecasts of future disease conditions in patients [5]. Diverse experiments were undertaken to predict various ML disorders, such as cardiovascular disease, and coronary heart disease. And the research focuses in particular on living COVID-19 predictions and the study Focuses even on the prediction and early response of COVID-19 outbreaks To monitor the current situation by decision making through the prediction process in order to direct earlier interventions in the efficient management of the disease.

The primary goal of the research is to provide the World Health Organisation (WHO) with an early prediction tool for the dissemination of new coronaviruses known as the SARS-CoV-2. At present threat to human life is COVID 19 Worldwide It has a spectrum from severe acute respiratory conditions and organ collapse and death in a very brief period[5]. It has many effects in the body of individuals. This worldwide pandemic impacts hundreds of thousands of people and causes thousands of deaths each day. There have been storeys of thousands of new people from all parts of the world being positive each day. The virus spreads primarily by contact with people, by breathable droplets or by touching the surfaces that have been contaminated. The most difficult aspect of its spread is that someone can have a virus without symptoms for many days. Nearly every country, on the grounds of its spread and its risk, The proclamation of the impacted areas and towns is partial or strict locks. Relevant vaccines and diseases drugs are currently being identified by medical scientists around the world. Since there are currently no licenced virus-destroying drugs available, the governments in all countries rely mostly on precautions to prevent the spread of the virus[10].Of all interventions, it is deemed highly necessary to be “conscious” of all facets of

* Corresponding author.
*E-mail address:* rameshkumar.mojjada@kluniversity.in (R.K. Mojjada).

https://doi.org/10.1016/j.matpr.2020.10.962
2214-7853/
Selection and peer-review under responsibility of the scientific committee of the Emerging Trends in Materials Science, Technology and Engineering.

COVID-19. A variety of scholars study the various aspects of the pandemic and generate findings to enable society to contribute to this aspect of knowledge.

Our objective is to develop a COVID-19 prediction system to contribute to the current human crisis. The three major variables for the next 10 days are forecast: 1) the no of new cases reported. 2) number of deaths 3) number of instalments of therapy. The analysis is based on some sophisticated supervised ML regression model (LR), the lowest absolute and selective shrinking operator (LASSO), vector supports (SMS) and exponential smoking (ES) models and has therefore proven to be a regression challenge in these circumstances. Education models have been trained with the data from patient statistics of Johns Hopkins' COVID-19. The data collection was prepared and subdivided into two sub-sets: preparation (85%) and assessment (15%). Output assessment for main measures, including R-squared value (R2 value), R-squared modified score (R2), MSD, MSD (MEA), and RMSE (Root Medium Square Error), were performed.

The following key findings are listed in this study:

- ES fits well if the time series data collection contains very minimal entries.
- Multiple ML algorithms tend to do best in different class predictions.
- A large amount of data is needed for predicting the future in most ML algorithms as data size improves model efficiency.
- .ML provision based model is very significant to the decision maker having pandemic like COVID 19

Six parts majority the text consists. Section I the methods and data set used in this study are described. The description is given in Section II. The results can be found in Section IV, the paper is summarised in Section V and closed. The findings can be found in Section IV.

## 2. Materials and methods

### *2.1. Dataset*

The goal of this research is to forecast COVID-19 'potential distribution with an emphasis the number of positive new events, mortality and recoveries. Data from the Git Hub[8,9] registry, supplied by Johns Hopkins University, Systems Science and Engineering Centre, were collected. The university mostly made the archive for the visual dashboard of the Novel Corona Virus in 2019 available and the ESRI Living Atlas Team helped it. A file called a Git Center repository (cssecovid19) includes data collection files.

### *2.2. Supervised machine learning models*

In the case that an undefined input instance is given, a supervised learning model can predict. Thus , data sets of input instances and its respective input instances for technique learning by the learning algorithms

The regressor is used for the regression model. A forecast for unpredictable entrants or test data is then generated in the qualified model. For the development of predictive models, Regression techniques and classification algorithms study method used here. In this COVID 19 prediction analysis , regression models of four are used:

- Linear Regression
- LASSO Regression
- Support Vector Machine
- Exponential Smoothing

A. Linear Regression

The aim class concentrates on individual regression simulation characteristics. It may also be used to define and model the relationship between independent variables and dependent. The most useful computer method for mathematical analysis of the machine learn is linear regression type regression simulation. A linear regression observation relies on two values, one on the dependence and one on the isolation. Linear Regression defines a linear relation between these variables' dependency and independence. Two variables (x , y) are necessary for the linear regression search. This equation indicates how y is associated with ×, which is called regression.

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{1}$$

$$E(y) = \beta_0 + \beta_1 x \tag{2}$$

This is the linear term for error regression. This error term takes into account the variability between × and y, β0 is the y-intercept and ß1 is the pitch.

A class mark is specified in the input data set for the purpose of the model × training of the linear regression in the context of machine study. The aim is to find the optimum values for β0 (intercept) and β1 (coefficient) to get the best regression line. The difference between the actual values and the values predicted should be minimum to make sure that this minimising problem is presented:

$$\text{Minimize}\frac{1}{n}\sum_{i=1}^{n}(pred_i - yi)^2 \tag{3}$$

here, g, which is the mean root square of the expected value for y (predi) and y (yi), n is the cumulative number of data points. g is called the cost function.

B. Lasso

LASSO is a regression model that is part of a linear regression method that uses shrinkage. Shrinking means reducing the extreme data sample values to the key values in this case. This strengthens and stabilises LASSO and reduces the error by the shrinking process. For multi-linear situations, LASSO is considered a more fitting model. LASSO thus makes the regression smoother in terms of the amount of functions it uses. It uses a form of regularisation to penalise additional tasks automatically.

However, the LASSO regression tries one at a time, because it does not add importance of zero if the new function would not boost the penalty term's fit with that function. The power of regularisation is therefore to automatically pick for us by adding the penalty for extra functions. Therefore, in this case of regularisation the models become sparse with few coefficients as the method removes values are zero. This regression LASSO acts to reduce the coefficient, which can be known by the square residual β slope), where, β slope is a concept of penalty.

C. Support Vector Machine

The SVM is a type of ML managed algorithm for reverse and regression classification. The SVM regression depends on a variety of statistical functions as a non parametric technician. The set of the kernel function converts data input into the form you like. In order to overcome regression problems using a linear function, SVM maps the vector(x) input(s) in the n-dimensional space called the function space(z) when dealing with non-linear regression problems. After linears regression is implemented in the space, non-linear mapping techniques are used for this mapping. Put the concept into an ML context using a number of observations with y from a multivariate training dataset (x) to N. The goal is therefore that the value of f (x) with (βrβ) as the minimum standard values is found as flat as possible. The dilemma then blends in with the minimization function. If the value of all residues is not greater than p, as in the following equation:

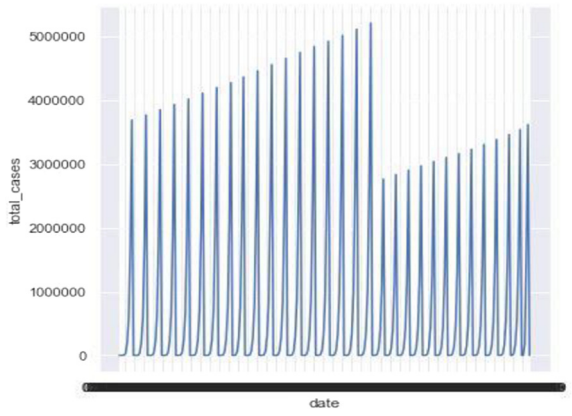


**Fig. 1.** Total number of cases.

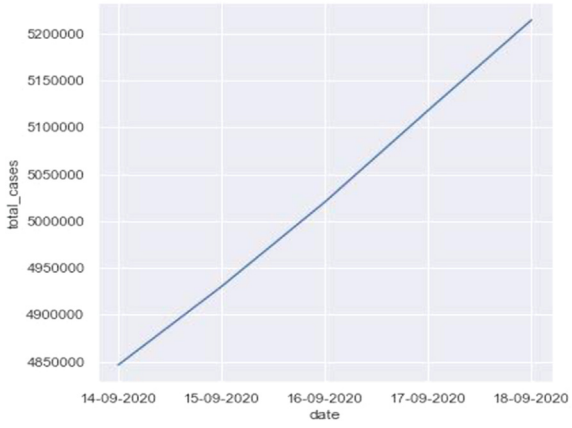


**Fig. 2.** Total cases of last 5 days.

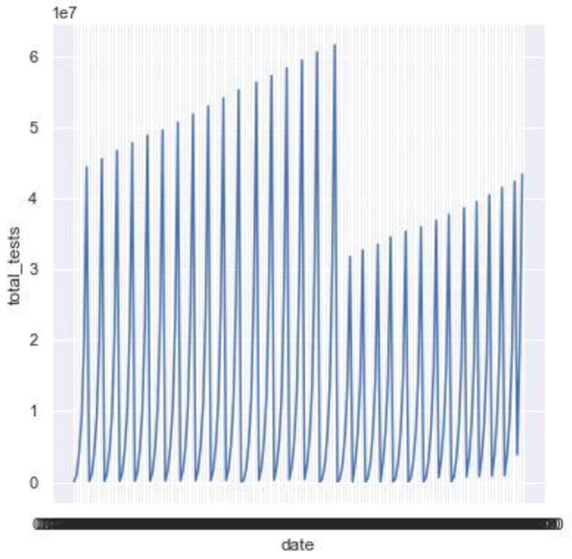


**Fig. 3.** Total Number of tests.

Predictions are rendered based on data from previous times of exponential family smoothing techniques. As previous data findings get older, their effect declines exponentially[11,12]. The weights are therefore geometrically reduced to the different lag values. Particularly for uni-variate is a time series provision. Ft 1 is the preview value of the previous prediction for the present period (Ft) in the ES. The prediction is as follows. Ft 1 is the real value

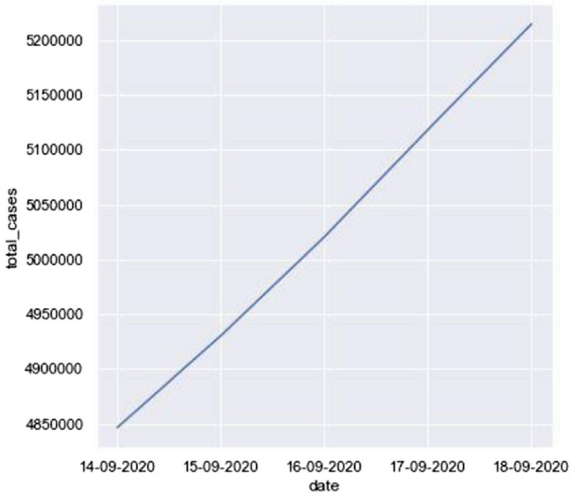


**Fig. 4.** Total tests in last 5 days.

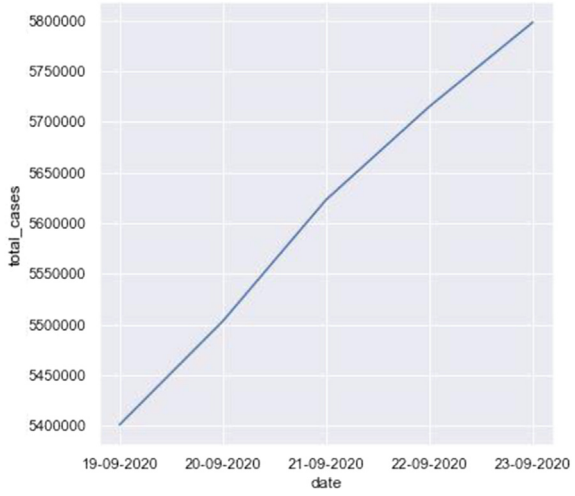


**Fig. 5.** Predicting total cases.

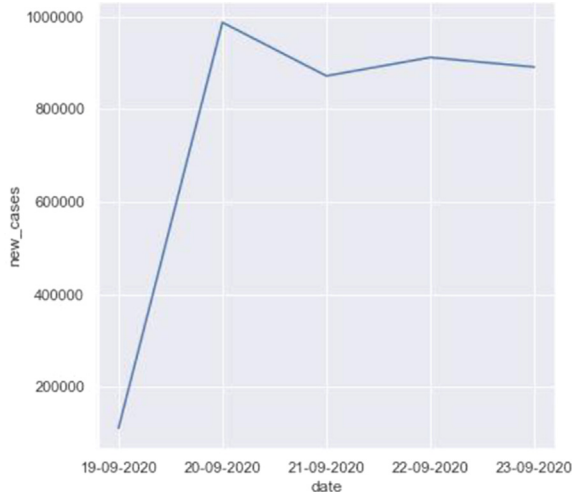


**Fig. 6.** Predicting total new cases.

in the preceding time frame, in which Ft 1 is the expected value in the prediction.

Evaluationparameters

The performance of each learning model is evaluated in the R2 scoring, Modified R-Square (R2), MSE, Mean Absolute Error (MAE) and Root Means Square Error (RMSE). In this analysis, the performance of every learning model is evaluated.

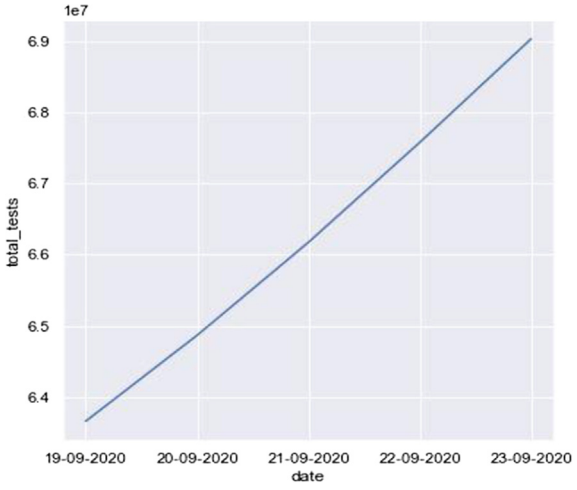


**Fig. 7.** Predicting total tests.

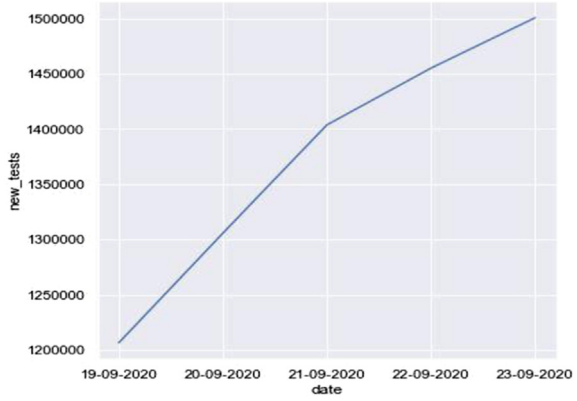


**Fig. 8.** Predicting new tests.

A. r-squared score

The R-square (R2) score for effective regression models is a statistical measure. These figures show the percentage of difference between the dependent and the independent variable. It is important to easily quantify 0–100 percent between the dependent differ and the regression model. We may verify the fitness of the trained models with the R2 scoring after the regression models. The R2 scoring reveals that the data points around the regression are spread, often referred to as the dissuasive component. Its ranking also varies between 0% and 100%.

The response variable is shown by 0 per cent as a result of the mean explained by the model, and 100 per cent as an average uncertainty of the answer variable. The high R2 value represents the consistency of the model learned. R2 is a linear formula describing the proportion of disparities.

B. Adjusted r-squared score

A transformed R2 shape is the Adjusted R-square (R2). This shows, like R2, the number of features in the forecast that the latter will adjust. The minimum normal value of f(x) with (βrβ). But this is the trend. For R2, the number of new features has grown. This will lead to an improvement if the new functions for the prediction model are helpful. However, once the new characteristics have been added, their value would plummet. The sample size is n and the sum of independent regression parameters is k.

C. Mean absolute error (mae)

This is the average of the test results between the model projections and the observed statistics of equal weight for all variations. The matrix range ranges from zero to endlessness, and fewer scores demonstrate the goodness of learning models, so it is also referred to as negative scores.

D. Mean square error (mse)

Another way of calculating regressive models output is a medium-square error. MSE takes and squares the regression line data points. Squared is significant if the negative sign of the value is omitted and greater weight is given to bigger variations. The lower the medium defect, the closer you find the better match. The better.

E. Root mean square error (rmse)

**Table 4**
Indian Sample Data.

| Day 1 confirmed cases | Day 100 confirmed cases | ……. | Day 262 confirmed cases |
|---|---|---|---|
| 0 | 540 | …… | 96,424 |

**Table 1**
COVID-19 patient death cases in India.

| Country | 31–12-2019 | 15–1-2020 | ……. | 04–09-2020 | ……. | 18–09-2020 |
|---|---|---|---|---|---|---|
| India | 0 | 0 | ……. | 68,472 | …… | 84,372 |

**Table 2**
COVID-19 new confirmed cases in India.

| Country | 31–12-2019 | 15–1-2020 | ……. | 04–09-2020 | ……. | 18–09-2020 |
|---|---|---|---|---|---|---|
| India | 0 | 0 | ……. | 83,341 | …… | 96,424 |

**Table 3**
Indian COVID_19 cases recovered.

| Country | 31–12-2019 | 15–1-2020 | ……. | 04–09-2020 | ……. | 18–09-2020 |
|---|---|---|---|---|---|---|
| India | 0 | 0 | ……. | | …… | 96,424 |

**Table 5**
Sample data of total death cases day wise.

| Category | 06–01-2020 | ….. | 04–04-2020 | …. | 18–09-2020 |
|---|---|---|---|---|---|
| Death | 0 | ….. | 68 | …. | 84,372 |
| Recovery | 0 | ….. | 235 | …. | 4,313,402 |
| New confirmed | 0 | …. | 601 | … | 96,424 |

**Table 6**
Day wise total new confirmed cases sample data.

| Day 1 deaths | Day 100 deaths | …… | Day 262 deaths |
|---|---|---|---|
| 0 | 177 | …… | 1174 |

A standard deviation from the expected error may be set to RMSE. The source applies to a square failure. forecast error are also notorious as residue as the space among finest match lines and real data points. Thus, RMSE tests the best fitness of the actual data points. RMSE is For the MSE square root, this is the following error rate.

## 3. Methodology

New coronoviruses, also called COVID-19 forecasts, are the main subject of the study. A potential human life hazard has been proved by COVID-19. Today's mortality rate rises across the world, leading to the death of tens of thousands. After the initial period of planning, the data collection have be split hooked on two sub-sets: a preparation for the model training and a test. This analysis aims to make possible mortal predictions. To track the pandemic, the effects of this analysis should be used. The thesis used apprenticeship models such as SVM, LR, LASSO, and ES. During days and recent reported cases, recovery [15] and trends of death these models were created. The rate of literacy, the number of infections registered each day, and the numeral of recovered cases over next ten days. Four acceptable ML approaches were required for this context. The empirical information offers summary tables on each day time-series, as well as the numeral of reported cases in recent days after the spread of the pandemic, deaths and retrievables [13.14]. At the beginning, the data collection of global data on daily numbers of deaths, confirmed cases and recoveries was processed for this study. The results[16] were derived from the recorded data displayed in Table 4 and the resulting data samples were shown respectively in Tables 5 and 6.

## 4. Results

### 4.1. Conclusion

A massive global catastrophe will arise from the precarity of the COVID-19 pandemic. Any experts and government organisations around the world are worried about the pandemic involving a significant percentage of the world 's population. In this study the possibility of a COVID-19 outbreak was suggested as a global ML-based predictive method. The framework analyses data sets containing real data from the past day to day and uses machine learning algorithms to make forecasts about future days. The study findings show that in view of the existence and dimension of the dataset, ES performs better in the current forecast domain. LR and LASSO have both been effective in estimating and verifying the death rate to some degree. The mortality rate will rise in the next few days and the survival rate will decline according to the effects of both models. The effect of ups and downs of the data collection is badly achieved by SVM in all the cases. A precise hyperplane between the data sets values was very difficult to place. In general, we assume that the present situation of the model forecasts are right and will help us understand the future. Therefore, the study projections will also allow the governments to act and take decisions in good time to avert the Crisis COVID-19. This analysis will be continually developed and we will analysis the process of the forecast using a modified dataset and apply ML methods of forecast with the most accuracy and appropriate. Echtzeit Live Projection will be a key focus of our future work. In this section, we shown the results of Covid-19 prediction using machine learning algorithm in terms of Figs. 1–8, and death, recovered cases of India mentioned in Tables 1–3.

## CRediT authorship contribution statement

**Ramesh Kumar Mojjada:** Conceptualization, Methodology, Software, Data curation. **Arvind Yadav:** Writing - original draft, Validation, Visualization. **A.V. Prabhu:** Investigation, Supervision, Software. **Yuvaraj Natarajan:** .

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

[1] P.W. Wang, F.G. Horby, G.F. Hayden, GaoA novel coronavirus outbreak of global health concern, Lancet 395 (10223) (2020) 470–473.
[2] S. Tuli, S. Tuli, G. Wander, P. Wander, S.S. Gill, S. Dustdar, R. Sakellariou, O. Rana, Next generation technologies for smart healthcare: challenges, vision, model, trends and future directions, Internet Technol. Lett.e145.
[3] B.V.V.S. Narayana K.S. Ravi N.V.K. Ramesh, (2018),'A review on advanced crop field monitoring system in agriculture field through top notch sensors', J. Adv. Res. Dyn. Contr. Syst.,10(6 Special Issue),PP. 1572- 1578
[4] S. Sridevi P.M. Bindu P.V.S.J. Krishna Teja , (2018),'User behavior analysis on agriculture mining system', Int. J. Eng. Technol. (UAE),7(2),PP. 37- 40
[5] Y. Bai, L. Yao, T. Wei, F. Tian, D.-Y. Jin, L. Chen, M. WangPresumed asymptomatic carrier transmission of COVID-19 JAMA (2020)
[6] D. Radhika K.D. Aruna The smart triad: Big data analytics, cloud computing and internet of things to shape the smart home, smart city, smart business & smart country International Journal of Recent Technology and Engineering, 8(2 Special Issue 11) 2019 PP.3594-3600
[7] Kalavala S.S., Sakhamuri S., Prasad B.B.V.S.V. (2019), 'An efficient classification model for plant disease detection', International Journal of Innovative Technology and Exploring Engineering, 8(7), PP.126-129.
[8] Likitha K., Aruna Sri P.S.G., Ratna Phanitha K. (2017),'A secure data sharing and revocation in cloud using IDE',International Journal of Applied Engineering Research,12(Special Issue 1),PP.228-235.
[9] S. Sakhamuri A. Virupakshi V. Pushpalatha D. Nagamani 'Elimination of redundant data in cloud with secured access control', International Journal of Innovative Technology and Exploring Engineering 8 6 2019 PP.1344-1348
[10] Y. Zhang, B. Jiang, J. Yuan, Y. TaoThe impact of social distancing and epicenter lockdown on the COVID-19 epidemic in mainland China: a data-driven SEIQR model study medRxiv (2020)

## Further Reading

[1] J.K.R. Sastry A.K. Sai R. Samuel D.B.K. Kamesh Architectural models for fault tolerance within clouds at infrastructure level' ARPN Journal of Engineering and Applied Sciences 12 11 2017 PP.3463-3469